Journal copyright restrictions and actual open access availability – a study of articles published in eight top information systems journals (2010-2014)

Mikael Laakso[1*], **Juho Lindman**[2]

[1] Information Systems Science, Department of Management and Organisation, Hanken School of Economics, Helsinki, Finland. e-mail: mikael.laakso, phone: +358 509100 864

[2] Applied Information Technology, University of Gothenburg/Chalmers, Gothenburg, Sweden

Abstract

Most scholarly journals have explicit copyright restrictions for authors outlining how published articles, or earlier manuscript versions of such articles, may be distributed on the open web. Empirical research on the development of open access (OA) is still scarce and methodologically fragmented, and research on the relationship between journal copyright restrictions and actual free online availability is non-existent. In this study the free availability of articles published in eight top journals within the field of Information Systems (IS) is analyzed by observing the availability of all articles published in the journals during 2010-2014 (1515 articles in total) through the use of Google and Google Scholar. The web locations and document versions of retrieved articles for up to three OA copies per published article were categorized manually. The web findings were contrasted to journal copyright information and augmented with citation data for each article. Around 60% of all published articles were found to have an OA copy available. The findings suggest that copyright restrictions weakly regulate actual author-side dissemination practice. The use of academic social networks (ASNs) for enabling online availability of research publications has grown increasingly popular, an avenue of research dissemination that most of the studied journal copyright agreements failed to explicitly accommodate.

Keywords

Open access, copyright, information systems, academic social networks

Introduction

The debate around open access (OA), i.e. free and unrestricted access to peer-reviewed research publications, has been ongoing for over two decades (e.g. Harnad 1995; Willinsky 2005; Suber 2012). There are diverging opinions on if and how OA should be integrated into the current practice of journal publishing which is still dominantly financed through subscription income (Beall 2013; Kingsley & Kennan 2015). The key arguments

in favor of OA are related to OA finally unlocking the power of electronic publishing; by providing free full-text access to readers OA has the potential to benefit the overall progress of science. For facilitating use and impact OA is likely to have benefits for individual authors and even journals in the form of wider readership and increased citations as most empirical studies so far indicate a positive correlation between OA and higher citation counts (Craig, Plume, McVeigh Pringle & Amin 2007; SPARC 2015). However, for journal publishers there is a risk in reduced subscription income should content be freely available instantly and comprehensively. Thus subscription journals usually restrict what distribution rights authors maintain to a manuscript which has been accepted for publication (Laakso 2014).

OA is usually split into two distinct types discerning between journal-provided access (gold OA) and author-provided access (green OA). The most prominent forms of gold OA are full OA journals and hybrid OA, both of which commonly involve author-side payment of a per-article article processing charge (APC), thus mitigating reduced or non-existent subscription income. Delayed OA is usually free for authors; it means that a subscription journal only requires paid access to read articles published for example within the last 12 months while all older content is OA (Laakso & Björk 2013). Promotional OA, which is also free to authors, can be argued to not be OA at all due to its temporary and sporadic nature. An example of promotional OA is when the publisher temporarily enables OA to sample issues or articles. However, as this study will showcase, leaving it out would omit an important part of the literature that is available to potential readers at any one time.

In contrast to OA articles that are made available directly on publisher websites in the form of the publishers final PDF, green OA is much more heterogeneous with regards to location and document contents. Green OA, also referred to as self-archiving, can be found across different types of web locations and the file provided can be anything from the authors initial draft submission manuscript to the publishers final PDF. 'Green' in this context comes from the notion of publishers giving a 'green light' for uploading openly available copies of the article contents. In some cases the article has not even been uploaded by the author but is made available on university course websites or seemingly random webpages. Many research organizations have a repository for making scholarly works of affiliated authors publicly available (opendoar.org), some research areas have discipline-specific repositories (oad.simmons.edu; Li, Thelwall & Kousha 2015), and researchers have been reported to be signing up to academic social networks (ASNs) in droves (van Noorden 2014; ResearchGate 2015a). ResearchGate, the largest ASN currently by number of hosted full-text documents, was founded in 2008 and has grown to over 7 million members in its 7 years of existence, with the five fastest growing research

disciplines on the service being medicine, engineering, biology, chemistry and computer science (ResearchGate 2015[a]). There are currently over 19 million full-texts available through the service (ResearchGate 2015[b]). Various different OA definitions exists (Laakso 2014), but Table 1 summarizes the key article document versions and web locations that help categorizing green OA. These same categories are used as-such for the study reported later in the article.

Table 1 Green OA (i.e. author-provided) document version and location categories. Based on earlier classification framework presented in Laakso (2014).

Document versions	**Definition**
Preprint	The article manuscript as submitted or prior to submission to the journal, prior to peer-review.
Accepted Manuscript	The authors' article manuscript as accepted for publication, after peer review but prior to the final copy-editing and layout of the journal.
Publisher Version	Version of record, publisher's version, published journal article.
Web locations	**Definition**
Institutional repository	Digital archives hosted by universities or research organizations that store content produced by affiliated authors systematically and persistently Although authors are the key contributors of content, professional librarians are usually involved in the quality assurance process, checking metadata and permissibility of upload as well as ensuring the long-term preservation of the content.
Subject repository	Open, non-commercial archives aimed at facilitating long-term storage and distribution of documents within specific scientific disciplines. Example repositories include ArXiv, Social Science Research Network (SSRN), and PubMed Central (PMC).
Academic social network	Web services where researchers can create personal profiles in order to communicate, collaborate and share content with each other. Some content might be restricted to members-only but a lot of uploaded content is also available to be indexed in web search engines and accessible by anyone on the open web.
Author website	Websites that are controlled by the author.
Other web site	Any website that does not fall into any above category.

In order for OA to be sustainable it has to adhere to the copyright restrictions and author rights defined by the journal that publishes the article. Failure to do so will risk the author or web service administrator receiving a cease and desist takedown notice by the journal publisher, particularly so if a single web location systematically distributes content of a certain journal in final published form. The most recent and illustrative example is Sci-Hub which has been sued by academic publishers for providing illegal access to millions of published scholarly journal articles (Bohannon 2016). Also well-established existing web services used by authors for self-archiving manuscripts are susceptible to changes in ownership, like the recent Social Science Research Network buyout by Elsevier (van Noorden 2016), which can bring with it a stronger degree of control for only allowing

copyright-compliant material uploads. The most comprehensive central database of journal publisher copyright agreements and author self-archiving rights is SHERPA/RoMEO (2016). An analysis of copyright agreements of the 100 largest scholarly journal publishers (by annual article volume) revealed that open distribution of the publisher's version is usually not permitted, but that over 80% of all published articles could be made available on institutional repositories as accepted manuscripts after 12 months of original publication (Laakso 2014). As the share of actually self-archived journal articles (or usually their manuscripts) has been measured in the range of 20-30% (Gargouri, Lariviere, Gingras, Carr & Harnad 2012), most scientific disciplines still have a lot of room to utilize fully free OA by researchers simply becoming more active in uploading their accepted manuscripts to web repositories.

Extant research

The most comprehensive and up-to-date estimate for the overall proportion of recently published (2011-2013) scientific journal literature available OA across all sciences pegged availability at nearly 54% of all published articles. The articles were found to be made available through a mix of content contributed by gold OA in journals and green OA distribution mechanisms (Archambault et al 2014). Two decades ago the OA availability of scientific articles was marginal at best. However, since then OA journal publishing and indirect author-side distribution have grown more sophisticated and professional. Archambault et al (2014) could discern that 12.1% of articles could be retrieved from full OA journals, 5.9% of articles from repositories, and 30.9% were retrievable by unknown mechanisms (e.g. delayed OA, hybrid OA, author homepages, random websites etc.). The study highlights the difficulty in accurately categorizing OA provisioning mechanisms in large sampling studies (Archambault et al (2014) included almost 514,000 articles out of a total 4.6 million Scopus article population) even when excluding manuscript version identification from the equation.

Jamali and Nabavi (2015) explored the availability of article full-texts published during the timeframe of 2004-2014, querying articles belonging to various Scopus research disciplines through Google Scholar with discipline-relevant keywords and recording data on the first page of query results. Of the 7244 articles queried for, 61.1 % (4426) contained a link to a full-text document. 80.8 % of these linked full-text articles were found to be the final publisher version. ResearchGate was the most frequent single source of full-text documents, accounting for 10.5% (466) of all full-text articles found.

Corroborating the results by Jamali and Nabavi (2015), ResearchGate was found to be a prominent source of free full-text access (only second after the US National Institute of Health) for a sample of 64,000 highly cited articles published during the timeframe of 1950-2013 (Martín-Martín, Orduña-Malea, Ayllón, & López-Cózar 2014). These two studies highlight the instrumental role that ResearchGate has come to play in providing full-text content for anyone to access through Google Scholar and other web search engines.

The most comprehensive previous study on the status of OA in Information Systems (IS) research specifically was conducted by Björk, Welling & Laakso (2011), which findings were also later reported in Björk & Paetau (2012). The study explored the availability of IS articles published in 2009, using Google to retrieve as many openly available copies corresponding to all articles published in eight top IS journals (IS basket of eight journals) (258 articles) as well as a stratified sample of articles from 44 other IS journals (540 articles). The study revealed a total OA share of 21.3%, with only 0.6% of articles being available on the journal websites directly and 20.7% on other web locations. The other web locations followed the distribution of: 8.3% subject-based repository, 32.7 % institutional repository, and 59% other types of websites. Of the documents found 38% were publisher versions, 46% accepted manuscripts, and 16% preprints.

The most recent insight into OA in IS research was provided by Lindman (2015), who conducted a small-scale empirical exercise investigating the OA availability of articles contained in the first issues of 2013 of eight leading subscription-based IS journals (IS basket of eight journals). Lindman's (2015) results showed that about half of the articles were available as OA in some form on the open web. We build on Lindman's (2015) results in this study, where we conduct a much more thorough search to map the current situation of OA in IS.

Methodology

This article aims to compose a comprehensive snapshot of the current relationship of actual OA availability compared to journal copyright restrictions, using the IS research context as a case. The methodology can be applied to any research discipline or cohort of journals as part of studies in the future.

This article investigates the OA situation in top IS journals concerning the following aspects:

1. What do the journals allow in terms of OA options (author-paid OA, journal self-archiving possibilities)?
2. To what degree is research published in the journals available openly on the web?
3. In what types of web locations can OA copies be found?
4. What versions of the articles are provided OA?
5. Do journal copyright statements related to self-archiving shape the OA availability of articles published in the journal?
6. How do citation counts compare between articles available OA and not-OA?

The IS basket of eight journals (European Journal of Information Systems, Information Systems Journal, Information Systems Research, Journal of AIS, Journal of Information Technology, Journal of MIS, Journal of Strategic Information Systems, MIS Quarterly) was selected to represent the journal population in this study. Originally published in 2007 as a list of six journals the IS basket of eight journals has been found to be largely representative of the IS field's overall diversity when it comes to representing research paradigms and methodologies (Myers & Liu 2009). Lowry et al (2013) conducted a bibliometric study focusing specifically on the IS basket of eight journals in an effort to retrieve empirical backing for the journals being selected as the top within the field. The conclusion based on the analysis of citation metrics what that the IS basket of eight can be identified as the top within the discipline also on empirical grounds. The IS field consist of many other journals, conferences, and publication outlets, but this methodological limitation is a practicality that is weighed against the implications of manual data collection. Narrowing the scope to eight key outlets, which have a high degree of homogeneity across them when it comes to topic focus, citation-indicators of quality, and authorship demographic enables a focused investigation on a well-defined area of scholarly communication within the IS discipline. While the journals are similar in the aforementioned ways, they still offer a beneficial variety in publishers where both large multi-journal publishers and smaller single-journal publishers are included. This delimited body of outlets, which have been in focus in several previous studies, enables a good balance in scope to explore the relationship between and copyright restrictions and OA availability.

There is no pre-existing database or centralized way of finding out which articles of specific journals have OA versions available on the open web, thus we opted for designing a manual data retrieval process. This replicates the circumstances of a potential reader interested in retrieving a certain published journal article. The research methodology design of Björk, Welling & Laakso (2011) served as the outgoing template that was iterated upon in the design of this study, with the aim of improving the level of insight on OA in IS provided by the results. The major methodological limitation of Björk, Welling & Laakso (2011) was that the study only recorded one OA copy per published article, the first one that was found to be available among the web search engine results. While one copy is usually enough for someone looking for a freely readable copy of a published journal article, it only paints a partial picture of the full breadth of OA availability since there are commonly more than one document floating around at any time across various kinds of web locations. The search for articles in Björk, Welling & Laakso (2011) was also limited to searching by title from google.com, not systematically incorporating a separate search in scholar.google.com which has rapidly grown to be a popular interface to finding and retrieving scholarly publications among researchers. Categorization of websites outside of systematic repositories was also coarse, where author personal websites and ASNs were not identified as separate categories. Finally, the study reported no journal-level metrics, instead focusing firmly on the discipline as a whole. Only an aggregate comparison of the total of the IS basket of eight journals versus all other journals was provided, resulting in that no discernible difference was found in the OA shares of the IS basket of eight journals compared to the sample of articles from other journals. These limitations informed the design process for the current study.

The research process can be divided into four major sequential stages:

1. summarizing and comparing journal copyright policies of the IS basket of eight journals,
2. acquiring the article and journal metadata for the entire population of published articles (IS basket of eight journals, articles published during the timeframe of 2010-2014),
3. searching for whether the identified article records have full-text copies openly available for download on the web, logging information for each found copy,
4. analysis and profiling of OA content, comparing journals to each other, document versions across different web locations, findings to journal copyright restrictions, citations to articles available OA vs. articles not OA.

Overview of current journal copyright policies

In order to provide an overview and comparison of what the copyright policies of the IS basket of eight journals in focus in this study permit and restrict when it comes to OA, Table 2 provides a summary of the most important factors compiled based on information available August 2015. Preprints usually have no outlined policy or very lenient restrictions. Accepted manuscripts can usually be distributed after an embargo of 12 or 18 months at specific types of web locations. It is only MISQ and JAIS that prohibit any distribution of the accepted manuscript. However, JAIS allows immediate distribution of the publisher version to the author's personal webpage while MISQ stipulates an embargo of 5 years from publication, the rest of the journals completely prohibit distribution of the publisher version. It can be concluded that the OA restrictions outlined here are not vastly differing from those of scientific journals in general across all sciences (Laakso 2014).

Table 2 Overview of the IS basket of eight journals, including copyright summary regarding OA policies

Journal	Publisher	JCR Impact Factor (2014)	Author rights to self-archive on the web (as per August 2015). When and where can authors make different versions of articles available OA?			Option of publishing OA in the journal against a fee (Hybrid OA)?
			Preprint	Accepted manuscript	Publisher version	
European Journal of Information Systems (EJIS)	Palgrave Macmillan	2.213	Anytime at the following locations: own personal website; employer's website; a free public pre-print server (e.g. RePEc)	18 months after publication, can upload to repositories.	Not allowed	Yes (2600 USD excl VAT)
Information Systems Journal (ISJ)	Wiley-Blackwell	1.766	Immediately upon submission to personal website and non-commercial repositories	12 months after publication to personal website, institutional repository, or specifically mentioned other repositories (inc. SSRN, ArXiV, RePEc)	Not allowed	Yes (3000 USD)
Information Systems Research (ISR)	The Institute for Operations Research and the Management Sciences (INFORMS)	2.436	Anytime to authors personal website, other locations must be disclosed prior to final acceptance	Immediately on personal website, after 12 months in non-commercial institutional repositories	Not allowed	Yes (3000 USD)
Journal of the Association for Information Systems (JAIS)	Association for Information Systems (AIS), Georgia State University	1.774	No policy	Not allowed, only publisher version allowed on personal webpage (immediately).	Immediately on personal, non-commercial web page	No
Journal of Information Technology (JIT)	Palgrave Macmillan	4.525	Anytime at the following locations: own personal website; employer's website; a free public pre-print server (e.g. RePEc)	18 months after publication to repositories.	Not allowed	Yes (2600 USD excl. tax)
Journal of Management Information Systems (JMIS)	M E Sharpe Inc (Taylor & Francis Group)	2.062	No policy	Immediately on authors personal homepage. 18 months after publication to repository or academic social network.	Not allowed	Yes (2950 USD excl. tax)
Journal of Strategic Information Systems (JSIS)	Elsevier	2.692	Anywhere at any time	Immediately on authors personal (non-commercial).homepage, or by updating existing pre-print in ArXiv or RePec. After 36 months on non-commercial hosting platforms (i.e. institutional repository), or commercial cites with which Elsevier has an agreement	Not allowed	Yes (1800 USD excl. tax)
MIS Quarterly (MISQ)	Management Information Systems Research Center, Carlson School of Management, University of Minnesota	5.311	No policy	Not allowed	5 years after publication on institutional repository, no other locations mentioned	No

Copyright information sourced from the following locations

EJIS (http://www.palgrave-journals.com/ejis/open_access_faqs.html) (http://www.palgrave-journals.com/pal/authors/rights_and_permissions.html#Self-archiving-policy)
ISJ (https://authorservices.wiley.com/bauthor/onlineopen_order_process.asp?jissn=1350-1917) (http://exchanges.wiley.com/authors/self-archiving_361.html)
ISR (https://www.informs.org/Find-Research-Publications/INFORMS-Journals/Author-Portal/Publications-Policies/Rights-Permissions) (https://www.informs.org/Find-Research-Publications/INFORMS-Journals/Author-Portal/INFORMS-Open-Option) (https://www.informs.org/content/download/43125/424945/file/copyright-130627-ISRE2.pdf)
JAIS (http://www.cb.wsu.edu/~tbeaulieu/JAIS/SampleCopyright.pdf)
JIT (http://www.palgrave-journals.com/jit/author_instructions.html) **(**http://www.palgrave-journals.com/pal/authors/rights_and_permissions.html#Self-archiving-policy)
JMIS (http://www.tandfonline.com/page/openaccess/openselect) (http://journalauthors.tandf.co.uk/copyright/Green-OA-AAM-embargo-periods.pdf) (http://www.tandfonline.com/page/openaccess)
JSIS (http://www.elsevier.com/journals/the-journal-of-strategic-information-systems/0963-8687/guide-for-authors#13700) (http://www.elsevier.com/about/company-information/policies/sharing)
MISQ (http://www.misq.org/permissions/)

Acquiring article titles and metadata (Scopus)

In the first stage a list of all articles published in the IS basket of eight journals for the years 2010-2014 was compiled. Scopus was used to search and retrieve bibliometric metadata for all published articles. Scopus metadata includes basic information about the articles such as title, journal, authors, affiliations, but also constantly updated metrics like received citations. The full metadata was downloaded for all articles on the 8th of February 2015, only excluding erratum and editorials from the Scopus exclusion criteria regarding document types. This resulted in total of 1422 articles. Despite Scopus being widely considered the most inclusive and comprehensive bibliometric indexing service it later became apparent that Scopus classification of article type is not perfect and the sample included some editorials, however, they were included in the sample in order to retain reproducibility in the methodology. From previous bibliometric research we were aware that Scopus indexing is not instantaneous with article publication and thus the integrity of the 2014 publication volumes were verified and corrected manually. 93 additional articles were added by comparing the extracted Scopus data and journal web page information on the published articles. As such citation information was unavailable for these titles, but that largely goes for all articles published in 2014 since few articles get cited within the first year of publication.

Searching for OA copies (Google and Google Scholar)

In the second stage each article was manually queried both through Google and Google Scholar to determine if an OA copy was available on the web (recording both manuscript version and web location for up to three separate copies per published article). The title of each article was used as the search string. The title was also used as the identifier for detecting any found OA copies; thus the analysis only acknowledges articles with identical titles. This methodological choice means that early self-archived versions with different titles might not be accounted for in the study, however, they are also often not a viable substitute for someone looking for the final published journal article. A similar approach is likely to be used by an interested reader trying to find an article based on e.g. a reference list entry. This search was conducted outside all university subscriptions to ensure that the contents would be delivered openly and not provided, and paid for, by a university library. Two researchers (the authors) were responsible for the full information search, coding, and analysis. The workload was divided roughly evenly so that both researchers conducted about 50% of information search for each individual journal.

Utilizing Google for ultimately defining what is available OA and what is not has both benefits and limitations. The benefit is that Google and Google Scholar are popular tools for students and academics, thus the data collection is grounded in the everyday practice of information seeking for many individuals within and outside of academia. The drawback is that Google is assumed to be aware of each and every document uploaded to the web, which has been studied and proven to be far from reality (Orduña-Malea & Delgado López-Cózar 2015). However, in lieu of using multiple search engines in parallel or scouring document repositories individually there is no obvious way to alleviate this limitation.

Analysis of the results

After searching for any available OA copies corresponding to the articles published in the journals, the retrieved OA articles were immediately categorized based on their web location and document version. The categories for describing the location were: 1) publisher's own web page, 2) institutional repository, 3) subject repository, 4) academic social network, 5) author website, 6) other web site. An additional categorization was later added for the articles of two journals: JAIS and MISQ, because it was not easily determined whether the freely downloadable PDF articles at AISNET (Association for Information Systems, aisnet.org) linked from Google were made free to access on purpose. As AISNET systematically provided a large amount OA of copies which were downloadable through PDF links among the Google search results they are separated since their future availability is uncertain. All results thus include separate calculations for including or excluding AISNET copies. As per the definitions given in Table 1 earlier, OA copy document versions were classified into either 1) preprint, 2) accepted manuscript, 3) publisher version, or 4) unknown version if there was no way to discern the version based on the PDF or information on the linked webpage.

Results and discussion

This section presents the research findings produced as a result of analyzing the collected data.

Overall results

Figure 1 provides an overview of the journal-level OA shares, i.e. the degree to which OA copies could be found corresponding to published articles in each individual journal. All article versions and web locations are

included in the calculations. The OA shares range from 32% (JIT) to 63% (ISR) of published articles. Including AISNET retrieved copies both JAIS and MISQ reached close to full OA coverage (98% and 92% respectively), however, even without AISNET the journals still had the second and third highest relative shares of OA articles available through other web locations (62% and 60% respectively).

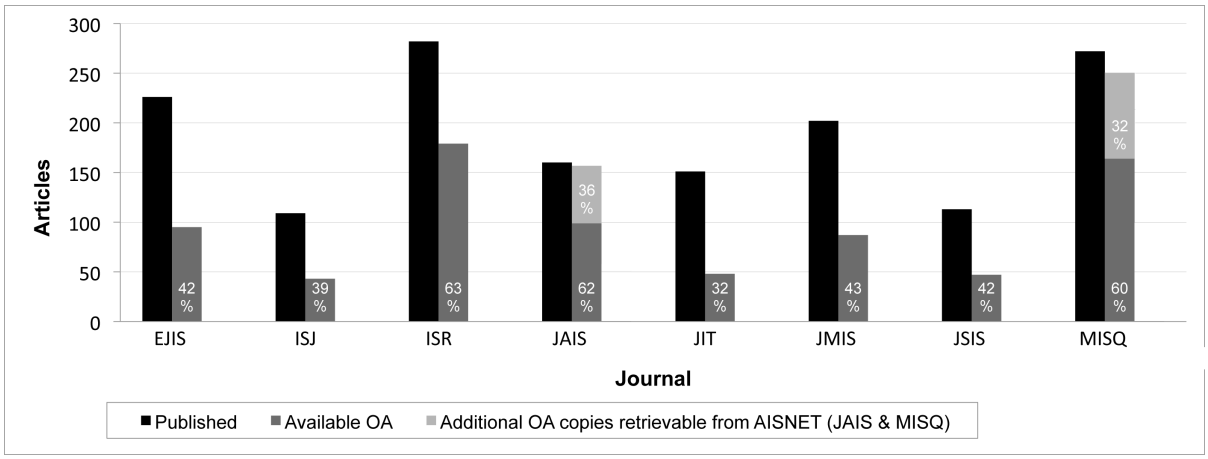

Fig. 1 Journal-level OA shares for articles published 2010-2014

Continuing with the journal-level metrics, Figure 2 presents a comparison of the relative frequencies of web location categories separately for each journal. Since each article can have up to three OA copies/observations attached this result provides a very holistic view of where top IS research can be accessed for free on the web. Even disregarding the massive AISNET influence on MISQ and JAIS the fluctuation of relative category shares between journals is notable.

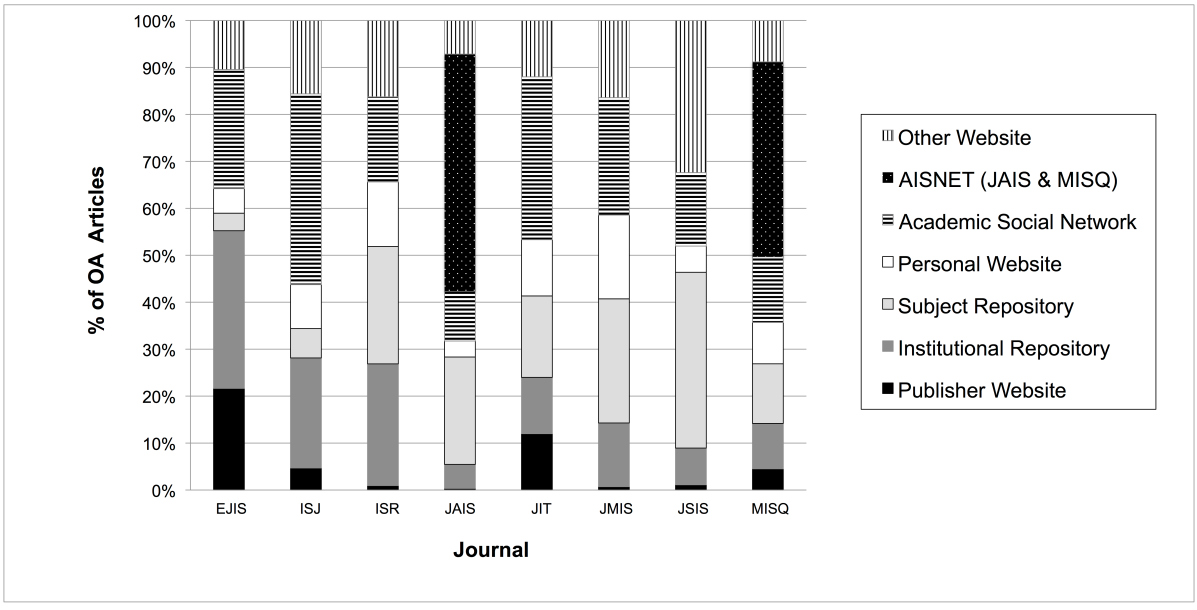

Fig. 2 Journal-level web location analysis for all OA copies found

Shifting the analytical lens to focus on what kind of article versions were downloadable from what kind of web locations, Figure 3 displays some interesting insights. Across all web location categories the publisher version was the most frequently appearing throughout.

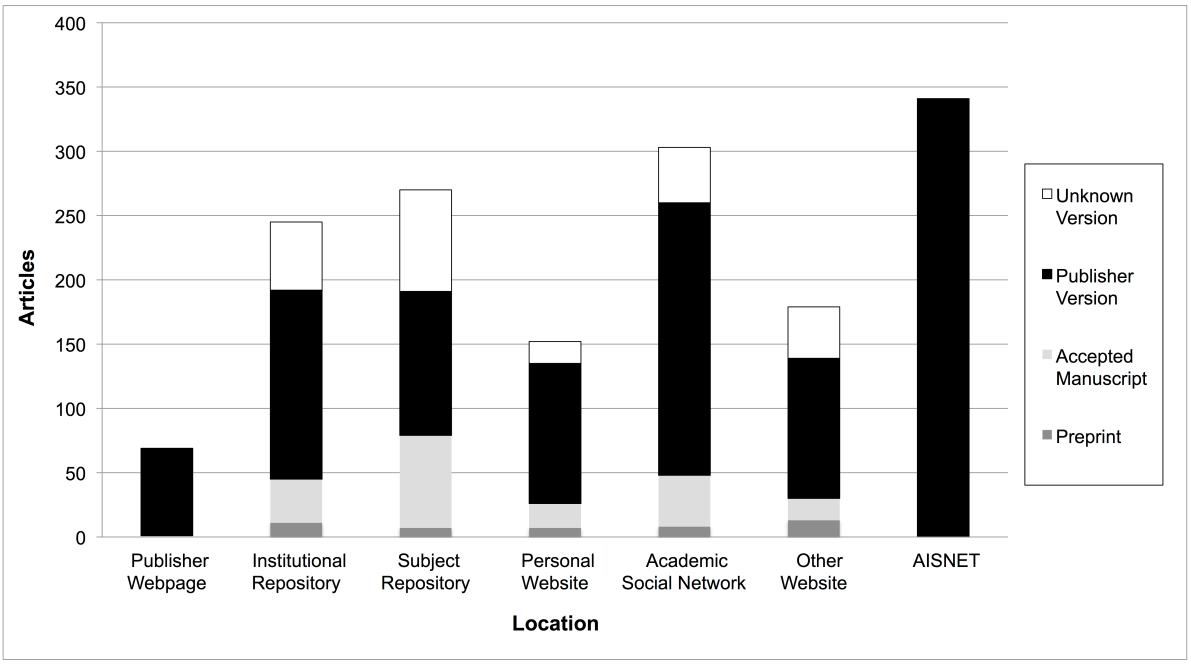

Fig. 3 Article versions across web locations for all OA copies found

Table 3 provides access to some of the key results in table form split by publication year, document version, and web location. While the study is not capable of offering insight into the longitudinal development of OA concerning these publications, their status as of spring 2015 showcases a slight increase in OA % for articles published between 2010-2012 (62% and 63%) in comparison to articles published in 2013-2014 (53% and 59%). This might be explained by many factors but some of the more likely include the additional time authors have had to get their work archived or publisher self-archiving embargos expiring which enables institutional repositories adhering to copyright restrictions to provide access. Nevertheless, the difference between publication years is small and supports the notion that OA copies are uploaded close to publication if they are uploaded at all. This arguably small variance in time from publication to OA, and the dominance of publisher versions from all journals being the most dominant document type across all web locations would suggest that journal copyright agreements are not effective at regulating article distribution. The overall OA share of around 60% for all articles (inc. AISNET) was found to be higher than the current best estimate available regarding OA across all sciences, which was recently pegged at around 54% (Archambault et al 2014).

Table 3 Overall results across articles published across all eight journals

Total	**2010**	**2011**	**2012**	**2013**	**2014**	**TOTAL**
Published articles	297	307	331	303	277	**1515**
Articles available OA	188	190	204	161	164	**907**
OA %	63%	62%	62%	53%	59%	**60%**
OA Articles found only on AISNET	22	35	33	24	31	**145**
OA Articles (AISNET copies not counted)	166	155	171	137	133	**762**
OA % (AISNET copies not counted)	56%	50%	52%	45%	48%	**50%**
Location breakdown						
Publisher webpage	**13**	**11**	**8**	**14**	**23**	**69**
Preprint	0	0	0	0	0	0
Accepted Version	0	0	0	1	0	1
Publisher Version	13	11	8	13	23	68
Unknown Version	0	0	0	0	0	0
Institutional repository	**64**	**44**	**61**	**42**	**34**	**245**
Preprint	5	2	2	0	2	11
Accepted Version	9	7	6	7	5	34
Publisher Version	36	24	44	25	18	147
Unknown Version	14	11	9	10	9	53
Subject repository	**54**	**54**	**71**	**51**	**40**	**270**
Preprint	1	5	0	0	1	7
Accepted Version	9	17	17	16	13	72
Publisher Version	37	19	41	11	4	112
Unknown Version	7	13	13	24	22	79
Personal webpage	**36**	**27**	**35**	**35**	**19**	**152**
Preprint	2	1	1	3	0	7
Accepted Version	2	4	3	7	3	19
Publisher Version	31	21	26	20	11	109
Unknown Version	1	1	5	5	5	17
Academic Social Network	**76**	**78**	**53**	**46**	**50**	**303**
Preprint	6	0	0	1	1	8
Accepted Version	7	10	8	7	8	40
Publisher Version	57	58	35	31	31	212
Unknown Version	6	10	10	7	10	43
Other website	**43**	**44**	**42**	**23**	**28**	**180**
Preprint	4	3	1	4	1	13
Accepted Version	6	1	3	3	4	17
Publisher Version	28	36	27	8	10	109
Unknown Version	4	4	11	8	13	40
AISNET (only JAIS & MISQ)	**63**	**71**	**85**	**51**	**71**	**341**
Preprint	0	0	0	0	0	0
Accepted Version	0	0	0	0	0	0
Publisher Version	63	71	85	51	71	341
Unknown Version	0	0	0	0	0	0

Access volatility and web domain analysis

Table 3 provides the full spectrum of OA article availability as recorded by the data collection, however, as one published article can be represented with up to three OA copies the insight into which web locations are the most critical, or volatile, for providing OA is limited. As such Table 4 provides a supplementary perspective on the data, this time considering which published articles had an OA copy only available at one single type of web location. From the analysis it is evident that ASNs contribute a lot of unique access to OA copies (92 articles/121 articles not counting AISNET copies) with institutional repositories coming in second place (77 articles/91 articles not counting AISNET copies). Institutional repositories should provide some of the most permanent archival of documents available, however, the future of widespread OA access through academic networks is still a large unknown factor.

Table 4 OA copies found only on one type of web location

										AISNET OA COPIES NOT COUNTED		
	EJIS	ISJ	ISR	JAIS	JIT	JMIS	JSIS	MISQ	**ALL**	JAIS	MISQ	**ADJUSTED ALL**
Publisher website	22	1	0	0	8	0	1	11	**43**	0	13	**45**
Institutional repository	24	6	30	0	4	4	6	3	**77**	7	10	**91**
Subject repository	0	1	17	1	1	10	1	1	**32**	36	9	**75**
Personal webpage	2	2	10	0	2	10	2	3	**31**	2	12	**42**
Academic social network	19	13	15	0	10	15	15	5	**92**	11	23	**121**
Other website	5	4	17	0	4	6	6	1	**43**	9	12	**63**
AISNET				58				89	**147**			

Most of all OA encountered was so-called green OA, where access to articles was provided through some other means than through the publisher of the journal directly. All 8 journals are subscription-based and thus do not provide their full content online for free like full OA journals. However, 69 articles were found to be available directly from the publisher webpages for one reason or another. Of these articles 29 were made available by EJIS and 21 by MISQ, with the other journals only having a few available each. These articles represent what the OA discourse would currently term gold OA. However, despite the glorified label, access to the identified copies is problematic. Most of them have likely been made free to download temporarily on a promotional basis or been retrievable directly through a link in the search engine due to a technical error. Hybrid OA, where articles are persistently available OA in subscription-based journals facilitated by author-side payment, is still a rare occurrence and its proper identification is challenging. One article identified as such is by Arvidsson,

Holmström and Lyytinen (2014) and was published in JSIS under a Creative Commons license. It is the only article from the 2014 volume of the journal to be available without a subscription.

OA copies found on author websites and institutional repositories are highly fragmented across different web domains, however, subject repositories and ASNs provide centralized hubs for researchers around the world to share their works on. Table 5 provides a closer look at the most frequent web domains and document versions for each category. The Social Science Research Network (ssrn.com) is clearly the most popular location with 173 articles found on the domain, with CiteSeerX (citeseerx.ist.psu.edu) coming in at second place with 93 articles. All three remaining domains (arxiv.org; mpra.ub.uni-muenchen.de, repec.org) accounted for a total of 4 articles. Subject repositories have a fairly even relative split for article version among accepted manuscripts, publisher versions, and unknown versions. Here the importance of authors properly marking uploaded copies is highlighted, unknown versions should either be clearly labeled as preprints or accepted manuscripts to increase the value of such copies to interested readers.

Table 5 Domain analysis for OA copies provided through subject repositories and academic social networks

Subject Repositories	**Academic Social Networks**
270 articles	**303 articles**
Domains	**Domains**
(173) ssrn.com	(303) researchgate.net
(93) citeseerx.ist.psu.edu	
(2) arxiv.org	
(1) mpra.ub.uni-muenchen.de	
(1) repec.org	
Document Versions	**Document Versions**
3% Preprints	3% Preprints
27% Accepted Manuscript	13% Accepted Manuscript
41% Publisher Version	70% Publisher Version
29% Unknown Version	14% Unknown Version

Among ASNs ResearchGate was shown to be a very popular venue for IS authors to share their publications openly on the web. A total of 301 OA articles could be found on the service – more than any other web location category outside of AISNET which is not really comparable since is the publisher website of JAIS and collaborates with MISQ to distribute content systematically. Other similar services came up among the search results, notably Academia.edu. However, that service requires registration prior to allowing downloading PDFs. Even so, while the authors would estimate that there were below 50 such hits to Academia.edu where a PDF could not be downloaded, ResearchGate still remains in a league of its own. The article version distribution on ResearchGate is heavily skewed to publisher versions with over two thirds of all copies downloaded from the

service being such versions. The results echo strongly with those of Jamali & Nabavi (2015) where ResearchGate came out as the top source for providing full-text articles. This seemingly "wild west" is a concern for the future sustainability of OA provided through such services, publishers are likely to take action should the scale of sharing start to eat into profit margins.

Citation analysis

To shed light on the debated relationship between OA and citations, a citation comparison between articles available OA and not OA was performed on both the full article population as well as individually for each journal. Both median and average (mean) number of citations received per article were calculated (Data source Scopus, as per February 2015) for both categories are provided (OA copy found/OA copy not found). Table 6 contains the relationship across all eight journals results as well as the aggregated calculation across all articles. The number of recorded OA copies available per published article fluctuated between 1 and 3 for articles where at least one copy was found, however, for this analysis the data was reduced to a binary OA/non-OA without considering the additional dimension of the number of OA copies available potentially impacting citations. Supporting this methodological decision is a recent study on Google Scholar data that found no correlation between the number of additional OA versions available and citation counts (Pitol & De Groote 2014).

Comparing citation counts for articles published within each individual year, the articles where an OA copy was found had a higher median and average citation count than articles where an OA copy was not found, regardless if AISNET copies are included in the comparison or not. To highlight the most notable differences: For 2010, where citations have accrued for the longest the median citations were 11 for non-OA articles and 17 for OA articles, for 2011 6 median cites non-OA and 12 median cites OA, for 2012 4 median non-OA cites and 7 median OA cites. Based on this strong evidence we argue that there is a positive relationship between OA and higher citations counts in top IS journals. However, since the relationship between OA and citations is complex and can materialize in many different ways, we urge further studies with larger annual article samples to be conducted before drawing generalizable conclusions.

Table 6 Citation analysis

										AISNET OA COPIES EXCLUDED		
	EJIS	ISJ	ISR	JAIS	JIT	JMIS	JSIS	MISQ	ALL	JAIS	MISQ	ADJUSTED ALL
Total articles articles published	226	109	282	160	151	202	113	272	**1515**			
Articles with no OA copy available	131	66	103	3	103	115	66	21	**608**	61	89	734
Articles with at least one OA copy available	95	43	179	157	48	87	47	251	**907**	99	162	760
Average number of OA copies available per published article	0,59	0,69	1,13	1,83	0,50	0,69	0,61	1,71	**1,03**	0,91	1,00	0,81
2010												
OA												
Average number of received citations	16,7	29,3	25,6	17,3	23,7	17,9	21,9	64,9	**30,2**	19,2	71,2	28,5
Median number of received citations	11	17	20	11	15	12	19	62	**17**	13	64	17
STD	17,7	35,4	21,2	15,5	27,4	15,3	12,9	59,5	**31,53**	16,3	67,6	29,5
NON-OA												
Average number of received citations	11,6	18,1	21,8	-	10,9	17,1	14,8	43,0	**15,7**	8,2	53,8	20,3
Median number of received citations	10	13	11	-	3	11	11	38	**11**	8	56	11
STD	10,1	12,6	19,4	-	20,4	18,8	15,1	18,7	**12,2**	4,9	41,0	18,4
2011												
OA												
Average number of received citations	12,0	11,0	15,6	9,7	7,7	11,9	20,2	39,7	**20,2**	10,5	50,8	20,0
Median number of received citations	6	8	11	8	7	9	13	28	**12**	9	37	11
STD	17,7	7,9	13,4	7,1	4,5	8,8	18,8	40,0	**22,2**	7,1	47,6	21,7
NON-OA												
Average number of received citations	8,6	13,6	11,9	-	4,3	10,2	15,1	-	**10,1**	8,2	26,9	12,7
Median number of received citations	6	12	11	-	2	6	11	-	**6**	8	18	8
STD	8,0	8,8	8,8	-	5,9	10,3	12,7	-	**7,6**	6,9	22,6	11,2
2012												
OA												
Average number of received citations	7,4	5,6	7,7	7,1	7,4	6,0	12,4	20,7	**11,0**	8,0	28,3	11,5
Median number of received citations	6	6	6	5	5	4	9	9	**7**	5	11	7
STD	5,9	3,3	7,7	8,0	5,5	6,2	6,9	36,5	**16,2**	8,8	45,9	16,5
NON OA												
Average number of received citations	6,3	6,3	6,3	-	4,3	5,9	7,3	-	**5,8**	4,7	10,1	6,4
Median number of received citations	4	6	6	-	2	4	3	-	**4**	3	8	4
STD	6,0	2,7	5,1	-	7,1	5,8	8,4	-	**4,6**	4,3	7,5	5,2
2013												
OA												
Average number of received citations	6,3	5,9	5,7	3,6	2,4	1,2	2,9	6,8	**5**	5,5	7,1	5,2
Median number of received citations	4	3	3	2	2	1	3	5	**3**	4	6	4
STD	4,9	5,1	5,8	4,6	1,6	1,4	1,9	5,8	**4,4**	5,7	5,9	4,3
NON-OA												
Average number of received citations	2,3	3,1	3,6	2,7	1,9	2,6	2,4	8,9	**3,3**	2,0	7,4	3,4
Median number of received citations	2	3	2	3	1	2	1	3	**2**	2	3	2
STD	2,6	2,1	5,3	2,1	1,9	3,7	3,1	21,6	**5,3**	1,8	16,1	5,4
2014												
OA												
Average number of received citations	0,6	1,7	0,2	0,4	0,5	0,4	1,7	0,0	**0,4**	0,6	0,0	0,5
Median number of received citations	0	1	0	0	0	0	1	0	**0**	1	0	0
STD	0,8	1,6	0,6	0,6	0,9	0,6	1,8	0,0	**0,7**	0,6	0,0	0,7
NON-OA												
Average number of received citations	0,9	1,6	0,0	-	0,3	0,6	0,3	0,0	**0,5**	0,4	0,0	0,5
Median number of received citations	0	1	0	-	0	0	0	0	**0**	0	0	0
STD	1,1	0,8	0,0	-	0,4	1,0	0,6	0,0	**0,6**	0,6	0,0	0,6

Conclusions

The purpose of this article is to inform as well as to invoke further discussion among the research community on the ways that in-progress and already published research is shared with others, highlighting the potential future implications of utilizing different dissemination options. Overall the OA landscape among articles published in top IS journals is currently vibrant and active. It is clear that IS scholars are already utilizing many different types of web services to distribute their work. In 2011 the overall OA share among the IS basket of eight journals was measured to be 20.4% utilizing a similar but more limited methodology (Björk, Welling & Laakso

2011). The jump to an OA share of 60% as measured by this study is thus evidence of strong growth in OA practices within the last five years. However, in addition an actual increase in OA copies being uploaded to the web one cannot exclude the impact of improvements in search engine coverage and indexing influencing the degree to which content can be discovered, five years is a long time when it comes to web technologies. Nevertheless, the end-result is that a higher degree of research can now be accessed. Despite there being no full OA journal among the eight journals, the discovered strong self-archiving practices provide broad visibility for research produced in these journals also outside the subscription-only journal websites. The study discovered that more could be done to benefit the impact of published research, findings upon which recommendations are outlined in the remainder of this article.

What came as a surprise to the authors was the large quantity of articles found on ASNs, in this case exclusively ResearchGate. Similar findings have also recently been reported in other studies on OA availability across scientific disciplines in general (Jamali & Nabavi 2015). The fact that over two thirds of all the google-provided PDF links to ResearchGate led to the publisher version, despite journal copyright agreements prohibiting their upload to any location (exceptions being MISQ allowing upload to institutional repository after five years of publication, and JAIS allowing the publisher version to non-commercial web pages only) raises concerns for the future sustainability of these services if large quantities of copyright-breaching content is being distributed. Only JMIS specifically mentions ASNs in the author copyright restrictions, the rest leave them unmentioned leaving their permitted use for distributing content unclear.

SSRN came out as the most popular subject repository for the articles in scope of this study. The major drawback encountered with documents uploaded to the service was that versioning was not made explicitly clear; often there was uncertainty if the manuscript uploaded to the service was identical to the one published in the journal. The utility, and thus likely impact, of content uploaded to the service could easily be increased by authors more clearly marking what version of the published article the uploaded manuscript corresponds to. The second most popular subject repository, and the only other one with strong support was CiteSeerX. However, CiteSeerX's mode of operation is much different than SSRN in that it does not allow authors to upload and manage their own documents but rather the service populates its archives by crawling the open web in order to download and re-host relevant full-text content. As such it is not a subject repository in the traditional sense, though its function to the community is roughly the same by aggregating literature around a specified topic area.

The higher citation counts for OA articles compared to non-OA articles is an interesting finding which causes and implications are hard to isolate and generalize. There is the potential factor of self-selection in authors only uploading OA copies of one's best work, thus leading to a higher citation count for OA articles compared to the rest. However, all of the journals included in the study are among the most prestigious within the field so this factor can be assumed to be of less importance than if a broader set of publications would have been included. There is also no way of knowing that the articles were actually OA at the time when the citing authors were looking for the article on the web. Nevertheless, whatever the many reasons behind higher citation counts might be, the implications for authors is that providing OA is likely to help in increasing visibility and impact for one's work – particularly in the IS field where the research is often relevant to practitioners and researchers in related disciplines alike where subscription-access to journals might be limited. A recent study indicated that papers uploaded to Academia.edu gained more citations over 1, 3, and 5 year time periods, even compared to the control group of articles that were found to be freely available online but not uploaded to Academia.edu (Niyazov et al. 2016). Like many other studies on the citation influence of OA (e.g. Sotudeh, Ghasempour, & Yaghtin 2015), this study does not prove causality between OA and increased citations since the element of self-selection cannot be omitted. However, there is something to be said for how ASNs can facilitate visibility and discoverability, providing additional value compared to just making publications freely available elsewhere on the web like e.g. institutional repositories or personal websites. As Orduña-Malea's and Delgado López-Cózar's (2015) study highlights, being open does not automatically mean being discoverable.

Limitations and avenues for future research

The main limitations of this study are related to the snapshot nature of its methodology, the limited set of journals included, and reliance on Google for defining the extent OA. A longitudinal study on the evolution of OA over time would be valuable but would involve following up the status of articles during multiple-years. However, this study can act as one measuring point for creating a longitudinal dataset for future studies on OA in IS.

This study provides a time capsule representing the current status of OA and journal copyright agreements, specifically in the context of top IS research. In addition to informing everyday practice and a more evidence-based discussion about OA we hope that this topic area is returned to by other scholars exploring other disciplines in a similar manner in the future. Even just identically replicating the exact methodology of this

study at a later point will likely produce a very different outcome as researchers, journals, and research funders are changing the way OA is facilitated and regulated.

References

Archambault, E., Amyot, D., Deschamps, P., Nicol, A., Provencher, F., Rebout, L., & Roberge, G. (2014). Proportion of Open Access Papers Published in Peer-Reviewed Journals at the European and World Levels: 1996-2013. Science-Metrix Report. Produced for the European Commission DG Research & Innovation. http://science-metrix.com/files/science-metrix/publications/d_1.8_sm_ec_dg-rtd_proportion_oa_1996-2013_v11p.pdf. Accessed 1st of July 2016.

Arvidsson, V., Holmström, J., Lyytinen, K. (2014). Information systems use as strategy practice: A multi-dimensional view of strategic information system implementation and use. *The Journal of Strategic Information Systems*, 23(1), 45-61. doi:10.1016/j.jsis.2014.01.004.

Beall, J. (2013). The Open-Access Movement is Not Really about Open Access. *tripleC,* 11(2), 589–597. http://triplec.at/index.php/tripleC/article/view/525. Accessed 1st of July 2016.

Björk, B-C., Welling, P., Laakso, M. (2011). Open Accessibility to Information Systems Research Articles. In *Proceedings of the 2011 ECIS Conference*. Paper 149. http://aisel.aisnet.org/cgi/viewcontent.cgi?article=1148&context=ecis2011. Accessed 1st of July 2016.

Björk, B-C., Paetau, P. (2012). Open access to the scientific journal literature – status and challenges for the information systems community. *Bulletin of the Association for Information Science and Technology*, 38(5). 39–44. doi: 10.1002/bult.2012.1720380512.

Bohannon, B. (2016). Who's downloading pirated papers? Everyone. *Science*. 352(6285), 508-512. doi: 10.1126/science.352.6285.508.

Craig I., Plume, A., McVeigh, M., Pringle, J. & Amin, M. (2007). Do open access articles have greater citation impact? A review of the literature. *Journal of Informetrics*, 1(3), 239-248. doi:10.1016/j.joi.2007.04.001.

doaj.org – The Directory of Open Access Journals. http://www.doaj.org. Accessed 20th August 2015.

Gargouri, Y., Lariviére, V., Gingras, Y., & Harnad, S. (2012). Green and Gold Open Access Percentages and Growth, by discipline. In E. Archambault, Y. Gingras, & V. Lariviére (Eds.), Proceedings of 17th international conference on science and technology indicators. Montre´al: Science-Metrix and OST. http://sticonference.org/Proceedings/vol1/Gargouri_Green_285.pdf. Accessed 20th August 2015.

Harnad, S. (1995). Electronic Scholarly Publication: Quo Vadis? *Serials Review*, 21(1), 70–72. http://cogprints.ecs.soton.ac.uk/ archive/00001691/00/harnad95.quo.vadis.html.

Jamali, HR, Nabavi, M. (2015). Open access and sources of full-text articles in Google Scholar in different subject fields. *Scientometrics*, 105(3), 1635-1651. doi: 10.1007/s11192-015-1642-2

Kingsley, DA., Kennan, MA. (2015). Open Access: The Whipping Boy for Problems in Scholarly Publishing, *Communications of the Association for Information Systems*. 37(1), 329-350. http://aisel.aisnet.org/cais/vol37/iss1/14. Accessed 1st of July 2016.

Laakso, M., Björk, B.-C. (2013). Delayed open access: An overlooked high-impact category of openly available scientific literature. *Journal of the American Society for Information Science and Technology,* 64(7), 1323–1329. doi: 10.1002/asi.22856

Laakso, M. (2014). Green open access policies of scholarly journal publishers: a study of what, when, and where self-archiving is allowed. *Scientometrics*, 99 (2), 475-494. doi:10.1007/s11192-013-1205-3

Li, X., Thelwall, M., Kousha, K. (2015) The role of arXiv, RePEc, SSRN and PMC in formal scholarly communication. *Aslib Journal of Information Management*, 67(6), 614-635. doi: 10.1108/ajim-03-2015-0049

Lindman, J. (2015). Open Access: A Cause, but not the Cause. *Communications of the Association for Information Systems*, 37(1), 352-356. Available at: http://aisel.aisnet.org/cais/vol37/iss1/15. Accessed 20th August 2015.

Lowry, P.B., Moody, G., Gaskin, J., Galletta, D.F., Humpherys, Sean, Barlow, J., and Wilson, D. (2013). Evaluating Journal Quality and the Association for Information Systems (AIS) Senior Scholars' Journal Basket via Bibliometric Measures: Do Expert Journal Assessments Add Value? *MIS Quarterly*, 37(4), 993-1012. http://aisel.aisnet.org/cgi/viewcontent.cgi?article=3136&context=misq. Accessed 20th August 2015.

Martín-Martín, A., Orduña-Malea, E., Ayllón, J.M. & Delgado López-Cózar, E. (2014). Does Google Scholar contain all highly cited documents (1950-2013*)? EC3 Working Papers*, 19. http://arxiv.org/pdf/1410.8464v4.pdf. Accessed 1st of July 2016.

Myers, M., Liu, F. (2009) What Does The Best Is Research Look Like? An Analysis Of The AIS Basket Of Top Journals. *PACIS 2009 Proceedings*. Paper 61. http://aisel.aisnet.org/pacis2009/61/

Niyazov, Y., Vogel, C., Price, R., Lund, B., Judd, D., et al. (2016) Open Access Meets Discoverability: Citations to Articles Posted to Academia.edu. *PLoS ONE,* 11(2), e0148257. doi: 10.1371/journal.pone.0148257

oad.simmons.edu - Disciplinary repositories. http://oad.simmons.edu/oadwiki/Disciplinary_repositories. Accessed 27th of March 2016.

opendoar.org - The Directory of Open Access Repositories. http://www. opendoar.org. Accessed 20th August 2015.

Orduña-Malea, E., Delgado López-Cózar, E. (2015). The dark side of open access in Google and Google Scholar: the case of Latin-American repositories. *Scientometrics,* 102(1), 829-846. doi:10.1007/s11192-014-1369-5

Ortega, J.L. (2015). How is an academic social site populated? A demographic study of Google Scholar Citations population. *Scientometrics*, 104(1), 1–18. doi: 10.1007/s11192-015-1593-7

Pitol, S.P., De Groote, S.L. (2014). Google Scholar versions: do more versions of an article mean greater impact?. *Library Hi Tech*, 32(4), 594–611. doi:10.1108/LHT-05-2014-0039

ResearchGate (2015[a]). Celebrating seven million members and seven years of ResearchGate. https://www.researchgate.net/blog/post/celebrating-seven-million-members-and-seven-years-of-researchgate. Retrieved 13th August 2015.

ResearchGate (2015[b]). ResearchGate Fact Sheet. http://www.researchgate.net/aboutus.AboutUsPress.downloadFile.html?name=rg_fact_sheet.pdf. Retrieved 13th August 2015.

SHERPA/RoMEO (2016) SHERPA/RoMEO: Publisher copyright policies & self-archiving. http://www.sherpa.ac.uk/. Accessed 20th August 2015.

Solomon, D. J., & Björk, B.-C. (2012). A study of open access journals using article processing charges. *Journal of the American Society for Information Science and Technology*, 63(8), 1485–1495. doi:10.1002/asi.22673

Sotudeh, H., Ghasempour, Z., & Yaghtin, M. (2015). The citation advantage of author-pays model: The case of Springer and Elsevier OA journals. *Scientometrics*, 104(2), 581-608. doi:10.1007/s11192-015-1607-5.

Suber, P. (2012). *Open Access*. MIT Press. http://mitpress.mit.edu/books/open-access. Accessed 20th August 2015.

SPARC (2015) SPARC - The Open Access Citation Advantage: List of studies and results to date. http://sparceurope.org/oaca_table/. Accessed 27th of March 2016

van Noorden, R. (2014). Online collaboration: Scientists and the social network. *Nature*, 512, 126–129. doi:10.1038/512126a.

van Noorden, R. (2016). Social-sciences preprint server snapped up by publishing giant Elsevier. *Nature News*. doi:10.1038/nature.2016.19932

Willinsky, J. (2005). *The Access Principle: The Case for Open Access to Research and Scholarship*. MIT Press. http://mitpress.mit.edu/books/access-principle